Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 4 (2011) 1953–1963
Procedia Computer Science

eScience: Building our Body of Knowledge

Valerie Maxville

iVEC
Perth, Australia
Email: maxville@ivec.org

Abstract

This paper describes the need for an eScience BoK, particularly as a resource for educators. eScience is a term representing the computational technology and techniques utilised when undertaking research. As eScience matures, stakeholders, and particularly educators, can benefit from the clarity that a defined Body of Knowledge (BOK) can provide. The BOK would require domain-specific and technological aspects to be addressed. This paper describes a framework for a prototype BOK for eScience and discusses how the BOK can be used as a tool to drive education, outreach and infrastructure planning.

Keywords:

1. Introduction

In the synergy of technology and science, eScience[1] has the potential to reveal new insights and entire research fields in science. In [1], the authors provide a generic knowledge acquisition cycle (Figure 1). When the social interconnections within science are included in this diagram, it becomes more like the scientific knowledge ecosystem in Figure 2. This synergy and transfer of knowledge goes beyond the boundaries of each project or discipline. Already the supercomputing infrastructure provided for energy research has transferred science, technology and engineering capabilities to climate change, spread of pandemic disease and hurricane modelling [2].

A shift towards sharing research and government data is taking place around the world: "*The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud-computing technologies*" [3]. Publishers are also supporting the push for data sharing, with the editors of Elsevier discussing openly accessible raw data, social networking and search and discovery trends in a recent update [4]. This may be facilitated in part through publishing APIs for delivery of intelligent information [5].

This synergy also exposes the need to develop expertise across the continuum from the application domain to the various technologies employed. In some projects, this may be very specific to the research and the researcher themselves may possess the required knowledge. In larger groups, a team with combinations of domain and technology may be required. In an example of an eScience project team [6] identifies four roles with distinct, complementary skills: Principal Investigator, Co-Investigator, Computer Scientist and Computer Engineer. Extending this to eScience support, the aim would be to have creative, flexible staff with a blend of technical and domain knowledge relevant

[1] In Australia the term used is eResearch, and hence eResearch BoK

Selection and/or peer-review under responsibility of Prof. Mitsuhisa Sato and Prof. Satoshi Matsuoka
doi:10.1016/j.procs.2011.04.213

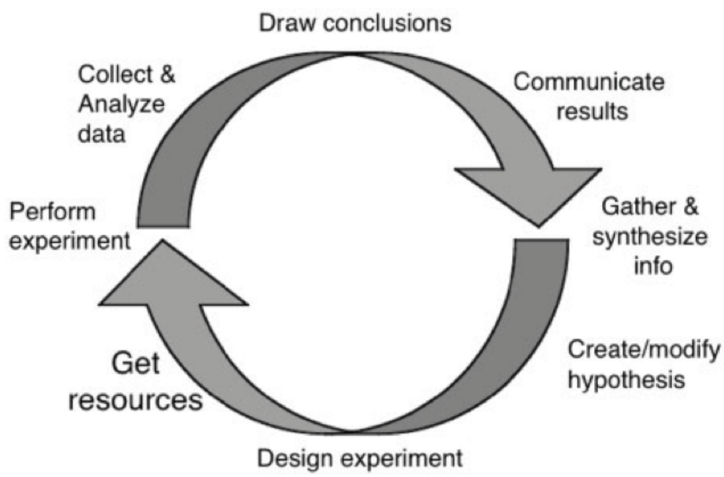

Figure 1: Scientific knowledge acquisition schema [1]

to a range of projects. The ICEAGE workshop report [7] indicates the types of students [graduates] required for eInfrastructure development are: Computer Scientists, Application Developers and Users and Systems Engineers and Managers.

Recognising the increasing demand for suitably trained R&D staff for eScience has led to symposia in the UK, US, and Australia. The Centre for Computational Thinking[2] considers how this awareness can be developed from the K-12 level. “Computational thinking is a way of solving problems, designing systems, and understanding human behavior that draws on concepts fundamental to computer science. To flourish in today’s world, computational thinking has to be a fundamental part of the way people think and understand the world.” Shodor[3] and the SC Education Program[4] are also working on aspects of this supply problem.

We now consider the challenges in eScience education and how the development of a shared knowledge base will support educators and researchers in eScience.

2. Education for eScience

The demand for computational thinkers has been identified as early as 1982, when the Panel on Large Scale Computing in Science and Engineering included “training of personnel in scientific and engineering computing” as one of the four components of a proposed National Program [9]. Ten years later, the need was again stated: “*Current curricula at grade schools and colleges will not educate students to exploit the possibilities opened up by parallel computers and the emergence of the computational methodology. [...] what we need most are **computational scientists** - individuals trained to **use** computers*” [8]. More recently, Anderson [10] quotes skills and knowledge shortages for general ICT practitioners across Europe, USA, Australia and Thailand due to very fast growth in ICT activity compared to enrolments. The US report on International Assessment of Research and Development in Simulation-Based Engineering and Science [11] found that

> Finding 2: Inadequate education and training of the next generation of computational scientists threatens global as well as U.S. growth of SBE&S. This is particularly urgent for the United States; unless we prepare researchers to develop and use the next generation of algorithms and computer architectures, we will not be able to exploit their game-changing capabilities. [11, p xv.]
> Finding 3: There is a clear and urgent need for a new, modern approach to educating and training the next

[2]http://www.cs.cmu.edu/~CompThink/
[3]Improving math and science education through the effective use of modeling and simulation technologies http://www.shodor.org/
[4]http://sc10.supercomputing.org/?pg=edprog.html

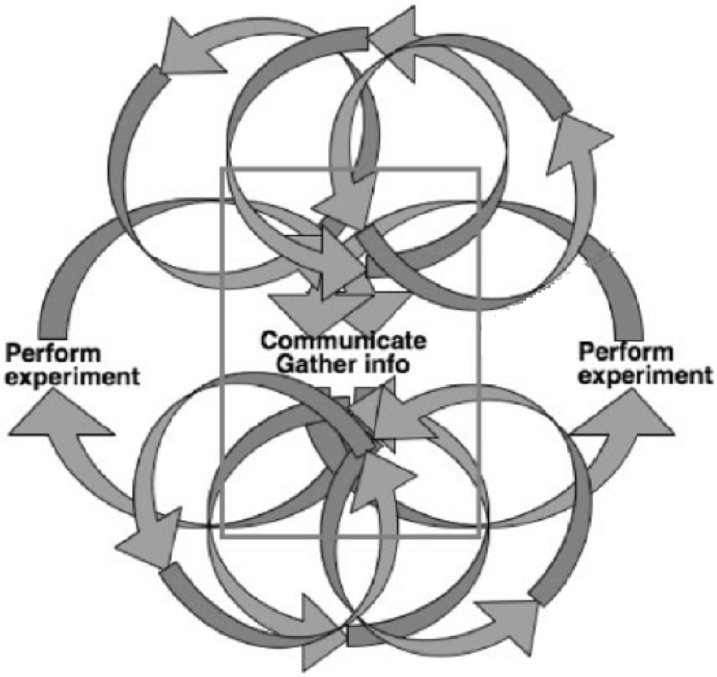

Figure 2: The scientific knowledge ecosystem [1]

> generation of researchers in high performance computing specifically, and in modeling and simulation generally, for scientific discovery and engineering innovation.[11, p xviii.]

A recent NNSA address highlighted the need for the technical base for creating computer simulations to provide solutions - predictive science. To have impact, we need to consider the desired outcomes for eScience education. The ICEAGE Curriculum Development workshop indicated the drivers as: "an urgent social and economic need to:
1) Equip first degree students in all disciplines with a level of skills in digital-systems judgement or computational thinking sufficient to support and progress the knowledge-based economy.
2) Invest in undergraduate and Masters courses to develop experts capable of innovating in the provision and exploitation of e-Infrastructures and e-Science" [7].

An additional challenge to education is the fluidity and rapid change in technology, along with the growing breadth of knowledge deemed relevant to eScience. The ICEAGE workshop refers to these challenges in developing curricula for e-Science as "*far from straightforward. Multiple methods and modes of delivery must be considered. Different target audiences would require the presentation of different principles, concepts and examples*"[7].

iVEC is based in Perth, Western Australia with the mission of fostering and promoting scientific and technological innovation through the provision of supercomputing and eResearch services. iVECs Education Program provides training to support use of its infrastructure. Our aim, however is to ensure that researchers have knowledge of the available techniques to allow them to select or develop appropriate tools and workflows to solve their research problems.

To support this, we believe that the development of an inclusively scoped Body of Knowledge (BoK) would be beneficial. The value of this would initially be in education and training - guiding course development, assisting in providing an overview of the fields to attendees of iVEC training and as extension material. The BoK could also be used be individuals for extending their skills and for career development. Researchers may find it useful for identifying technology applicable to their research and to help define the skills required for research teams. The process of building the BoK should assist in highlighting similarities across disciplines, for example, techniques used in materials science that are common between chemistry and physics.

3. Building a BoK

A Body of Knowledge can be defined as: (1) "Structured knowledge that is used by members of a discipline to guide their practice or work." (2) "The prescribed aggregation of knowledge in a particular area an individual is expected to have mastered to be considered or certified as a practitioner." [12]. Fields with published BoKs include:

- Software Engineering Body of Knowledge (SWEBOK)[5]
- Project Management (PMBOK)[6]
- The ICT Profession Body of Knowledge (CBOK)[7]
- Usability Body of Knowledge[8]

eScience pervades all other disciplines, thus using the BoK to set boundaries would not be productive. It does not displace but extends and interprets other BoKs from related areas.

A well-established BoK is the Software Engineering Body of Knowledge (SWEBOK), implemented as a Guide to the SWEBOK [13]. The project plan comprised three successive development phases: Strawman, Stoneman, and Ironman. An early prototype, Strawman, demonstrated how the project might be organised. The publication of the Trial Version of the Guide in 2001 marked the end of the Stoneman phase. Trial usage and feedback on the Ironman phase resulted in a Guide that has achieved community consensus.

The Guide to the Software Engineering Body of Knowledge (SWEBOK) was established with the following five objectives:

1. To promote a consistent view of software engineering worldwide
2. To clarify the place - and set the boundary - of software engineering with respect to other disciplines such as computer science, project management, computer engineering, and mathematics
3. To characterize the contents of the software engineering discipline
4. To provide a topical access to the Software Engineering Body of Knowledge
5. To provide a foundation for curriculum development and for individual certification and licensing material

All of these objectives would be considered of value for eScience.

Recording a BoK requires the definition of terms and main conceptual groupings for the field. This is coloured by the perspective taken, for example the view of an algorithm from computer science may differ to how it is viewed from the application domain. As the fields and application areas drive the science, and are the researcher starting point, this would be a preferred lead perspective. The technology and techniques used in the fields can be factored out and assembled to provide and overview from the technical perspective. This differs from the SWEBOK which builds on ten technical knowledge areas. There are additional topics that become important when undertaking eScience, such as data issues for ownership, licencing and ethics, which will need to be included.

Discussions of concepts and terms across disciplines/communities is often supported by ontologies. This defines the objects, classes, attributes and relations in an area. For example, in material science, objects may include types of problems and the techniques applied to them. A class may be used to define groupings of problems and techniques. Ontologies exist for genes[9] and proteins[10] and allow for the definition of terms and their relations. A visualisation of an ontology can help to communicate the concepts and how an area fits together.

The SWEBOK was developed through community input over a period of time. This was facilitated through documents for comment and workshops. Another common approach for recording a knowledge base is via wikis such as wikipedia[11]. The Usability BoK provides a draft BoK and wiki for community comment[12]. Subject specific

[5] http://www.swebok.org/
[6] http://www.projectsmart.co.uk/pmbok.html
[7] http://www.acs.org.au/attachments/ACSCBOKWorkingPaperV5.0Oct2008.pdf
[8] http://www.usabilitybok.org/
[9] The Gene Ontology: http://www.geneontology.org/
[10] PRotein Ontology (PRO): http://www.obofoundry.org/cgi-bin/detail.cgi?id=protein
[11] http://en.wikipedia.org/
[12] http://www.usabilitybok.org/

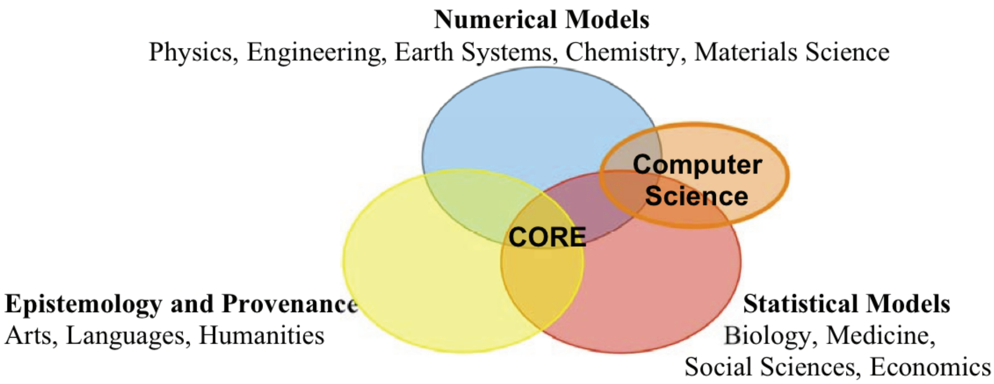

Figure 3: Dominant techniques by discipline, with Computer Science shown to include both numerical and statistical models [7]

wikis include the SWAN Project[13], a project to develop a knowledge base for the neurodegenerative disease research community. SWAN has an ontology ecosystem generated through Semantic Web technology and self-organised online communities. Wikis provide a transparent and open process for disseminating, tracking and reviewing documents and allow for entire communities to be editors.

The preferred approach in this case will be the development of a strawman (prototype) BoK from existing resources. Initially this will be housed on the iVEC Education web pages. This can then be made available and editable for the eScience community via a wiki or similar. Feedback will also be sought through Birds of a Feather sessions and workshops at conferences such as eScience, Supercomputing, ICCS and eResearch Australasia.

4. An eScience BoK

Many useful resources exist that can be drawn upon in building a BoK for eScience. The BoK itself can focus on high level coverage and refer to these materials. A BoK with detailed descriptions of all of the techniques would be difficult to maintain (with a rapidly changing field) and be likely to be unbalanced in depth of coverage.

As this initiative is coming from the education perspective, the existing curricula and surveys around eScience, supercomputing, data management, visualisation, collaboration and networks will be included. The knowledge areas will encompass the traditional disciplines as visualised in Figure 3. Although this diagram is helpful in communicating the commonalities between disciplines, the BoK will be more easily understood if opened out into a flat format and links added to represent interactions.

A more formal structure for the knowledge areas may be drawn from the categories defined for the Australian and New Zealand Standard Research Classification (ANZSRC)[14]. The ANZSRC was developed by the Australian Bureau of Statistics and Statistics New Zealand. It is used in applications and reporting on competitive grants in Australia, and is the basis for the subject element in the RIF-CS schema used by the Australian National Data Service (ANDS)[14], making it a suitable choice in the Australian context. The terminology used in the ANZSRC may provide a basis for an eScience ontology.

The ANZSRC Field of Research (FOR) classification is where the methodology used in the R&D is considered. The FOR has three hierarchical levels: Division, Group and Field. Table 1 provides an example of levels in the Physical Sciences. There are 22 Divisions (2-digit), 157 Groups (4-digits) and below them 1238 Fields (6-digit). The added value of this system is that where a field could be assigned to more than one category, the ANZSRC defines which area it is placed in/excluded from. This BoK will use the Division level (Table 2) to define the majority of the "Knowledge Areas" for the BoK.

In addition to these, the BoK will cover Technology and Techniques as two separate Areas. In characterising the Areas, the prototype BoK will use the following format: overview of KA, including sub-areas, then consider each

[13]Semantic Web Applications in Neuromedicine: `http://swan.mindinformatics.org/`
[14]`http://www.ands.org.au/resource/rif-cs.html`

Table 1: ANZSRC Field of Research for the Physical Sciences[14]

Level	Code	Description
Division	02	Physical Sciences
Group	0201	Astronomical and Space Sciences
	0202	Atomic, Molecular, Nuclear, Particle and Plasma Physics
	0203	Classical Physics
	0204	Condensed Matter Physics
	0205	Optical Physics
	0206	Quantum Physics
	0299	Other Physical Sciences
Field	020603	Quantum Information, Computation and Communication

Table 2: ANZSRC Field of Research (FOR) Divisions[14]

Code	Description	Code	Description
01	Mathematical Sciences	12	Built Environment and Design
02	Physical Sciences	13	Education
03	Chemical Sciences	14	Economics
04	Earth Sciences	15	Commerce, Management, Tourism and Services
05	Environmental Sciences	16	Studies in Human Society
06	Biological Sciences	17	Psychology and Cognitive Sciences
07	Agricultural and Veterinary Sciences	18	Law and Legal Studies
08	Information and Computing Sciences	19	Studies in Creative Arts and Writing
09	Engineering	20	Language, Communication and Culture
10	Technology	21	History and Archaeology
11	Medical and Health Sciences	22	Philosophy and Religious Studies

Phase of research and the techniques used in each. The phases align to the cycles of research in Figure 1. Some commonality is expected between sub-areas and across the FOR Knowledge Areas. It is expected that the phase with the most specific techniques will be in the experimentation/computation for each area. Table 3 maps the framework across the Knowledge Areas and Research Phases.

Table 3: eScience Body of Knowledge framework

Research Phase	Technology	Techniques	FOR Knowledge Areas
Data Collection			
Experimentation			
Analysis			
Communication			
Archive/Share			

4.1. Technology Knowledge Area

When introducing eScience, the iVEC approach has been to outline the five main technology resources provided by our organisation: Compute; Data; Visualisation; Networking and Collaborative tools.

Each of these areas is complex enough to have its own Body of Knowledge. The typical compute is via a supercomputer running parallel code. Alternatives might be distributed systems (for example Grid, Cloud, BOINC), hybrid systems and standalone workstation processing. Any of these may be used in the Research Phases for data collection, experimentation or analysis.

When working with Data, researchers can use their local storage and media, institutional or group servers, the iVEC petascale datastore or the ARCS[15] Data Fabric. A range of these would be used through the life of a research project.

Networks are the key to much of eScience. Aspects of interest would be latency, protocols and security issues. These are most likely to have an impact on collaboration tools and on moving data between machines.

Visualisation can take place on dedicated visualisation resources, or increasingly at researchers workstations. Dedicated options include front/rear projection, immersive systems, 3D stereo and high resolution display walls. Other user interface options are also grouped under visualisation, including haptics, motion and eye tracking, EEG[16] and user monitoring. Visualisation takes place mainly in the first four Research Phases (Table 3).

On the social side, collaborative tools can have a strong impact on the operation and interaction of teams. Collaborative tool can run the spectrum from chat to multi-cast video-conferencing, from email to wikis and portals. These can be classified in terms of their collocation and sync/async nature, as shown in Table 4.

Table 4: Collaboration Tools

	Same Location	Different Location
Different Time	Whiteboard, Noticeboard	Fax, Post, Email, Web, SMS, Wiki, Social Networking tools, Online Forums
Same Time	Meeting, Water-cooler conversation	Access Grid, Telephone, Videoconference, Teleconference, Instant Messaging

Shared instruments are also included within data collection and experimentation. These include Synchrotrons, the Large Hadron Collider, telescopes and sensor networks. An example of organised knowledge to support user access to instruments is the Australian Microscopy and Microanalysis Research Facility (AMMRF)[17]. Researchers can access a complete list of available resources, or work through the Technique Finder, entering type and scale for their investigation to view a list of recommended techniques.

As with all of the Knowledge Areas, the information in the BoK will always be representative, not exhaustive. Localised information on the technology available would be of value for educators and researchers.

4.2. Techniques Knowledge Area

The Techniques refer to eScience processes collated from use in the FOR Knowledge Areas. They can be considered in reference to the same categories as the Technologies, as well as by the Phase of the Research. The following discussion is a selection of the Techniques in the general, parallel, data management and visualisation areas.

There are also some techniques across the entire project, such as ethics, project management, and software engineering. Social aspects of teams collaborating across geographic, time zone and cultural barriers are common. Collaborative tools and organisation of virtual teams have well-developed knowledge bases to draw upon.

A previous paper on preparing scientists for scalable software development identified the key SWEBOK areas based on the size and audience of the project. Table 5 gives an overview of the SWEBOK Knowledge Areas and their priority as a skill-set for researchers/teams.

The ICEAGE Report [7] included the following topics in the Programming for e-Science component: loosely-coupled programming (includes communications, networks issues, workflows...); programming to APIs; Code re-use & component publishing; API production; code maintenance, versioning, etc.; technical documentation for re-use; standards; programming environments; security; introduction to existing CS methods & concepts.

One of the more established areas within eScience is parallel code development. The International Working Group on Software Engineering for Parallel Systems (SEPARS) publishes an overview of parallelism and multi-core/manycore in curricula around the world [16]. Their classification of course topics includes: algorithms; architec-

[15] Australian Research Collaboration Service: `https://www.arcs.org.au/index.php/arcs-data-fabric`

[16] Electroencephalography (EEG) is the recording of electrical activity along the scalp produced by the firing of neurons within the brain (wikipedia).

[17] `http://www.ammrf.org.au/`

Table 5: Project stage and Coverage of SWEBOK Knowledge Areas [15]

Knowledge Area	Individual	Group	Public
Requirements	++	+++	+++
Design	++	+++	+++
Construction	+	++	+++
Testing	++	+++	+++
Maintenance	+	++	+++
Configuration Mgmt	++	+++	+++
Engineering Mgmt	+	++	+++
Engineering Process	+	++	+++
Tools and Methods	++	+++	+++
Quality	+	++	+++

+ light ++ deeper +++ formal coverage

ture/hardware; programming; distributed computing; multicore programming; scientific computing/HPC; theory of parallel computing; and no classification.

In the practical nature of a BoK, this would need to be extended to include more specific topics such as languages, task and data parallelism, memory hierarchy, benchmarking, debugging and libraries. Platforms types are also expanding, and matching techniques are required for distributed computing, grid, cloud, GPU and heterogenous computing.

Many resources exist in software development for HPC, such as SEPARS, SC Education Program and VSCSE[18]. The last two are taking up a focus on petascale and exascale computing - a computing power transition that will require new software and tools at all levels of supercomputing. Algorithms, their implementation and their performance profiles are key Technique resources that are useful across all Knowledge Areas.

The relatively recent shift to open data in research has made data management a critical area. The Australian National Data Service (ANDS)[19] provides guides for research data management. When working with repositories, [6] highlights digital rights management, copyright ownership, database rights, fair dealing, licences and Z39.50 in the "Introduction to EGEE" course. Atkinson [**?**] provides an outline of a masters course in eScience with a data management component including: storage, movement, provenance, life-cycle, validation, security, schemas / data formats and documentation. The data track at eResearch Australasia adds to this list with : Data Commons; Data Grids and Clouds; Building Data Management Capabilities; Data Utilities; Generating Data; Discovering Data; Exchanging Data; Reusing Data; Combining Data; and Publishing Data.

Visualisation can be used to check input data, as a key result of experimentation/computation, for analysis and for communication of results. Information visualisation can utilise mind maps, conceptual diagrams, visual metaphors and concept mapping [17]. A popular analysis and visualisation tool is social network analysis and mapping techniques to non-social and non-geographic data, showcased in the Atlas of Science [20]. Bresciani [18] indicate that interactive and annotation-friendly visualisations have a positive impact on knowledge sharing, regardless of the form of the diagram.

Many computational science application areas consider visualisation critical, including Computational Fluid Dynamics and Material Science. In scientific visualisation, the data may be made up of points, a rectangular grid, a curvilinear grid or an unstructured grid [19]. These can be surface or volume rendered and the information enhanced through surface shading, isosurfaces, clipping/slicing or adding particle traces. For communication purposes, the relative merits of images (frames), movies/animations and interactive models needs to be considered.

[18] Virtual School of Computational Science and Engineering: `http://www.vscse.org/summerschool/2010/petascale.html`

[19] `http://ands.org.au`

[20] `http://scimaps.org/atlas/`

The Techniques Knowledge Area will continue to expend and for some may be the key Section of the BoK. However, for a scientist, the FOR Knowledge Areas will be most relevant. This will give a select set of techniques which can be traced through to broader information and then link to more detailed discussions.

4.3. FOR Knowledge Areas

There is wide variation in the uptake of eScience across the 22 divisional categories. It is likely that some groupings of these Knowledge Areas are possible to reduce repetition and to further identify commonality between fields. Candidates may be in the humanities and social science areas. Although, over time, there may be more new techniques that result in differentiation.

This section will highlight three of the courses in the SC Education Program 2008 and 2009. Some surveys of computational techniques exist, including biochemistry networks [20] and quantum dots and nanosystems [21]. These, and similar articles, will provide a resource for generating coverage of techniques. Summarising the topics in the SC training courses can provide a quick overview of established techniques and tools in fields. Table 6 provides this for Biology, Chemistry and Engineering.

Table 6: FOR Knowledge Area Summaries

Field	Techniques and Tools
Biology	Bioinformatics, Computational Genomics, Dynamic Modelling, Molecular Phylogenics, Transcriptomics, Homology Modelling Tools: MATLAB, BLAST, R
Chemistry	Molecular Mechanics, Ab Initio Hartree Fock, Semiempirical and Density Functional Theory Tools: SIESTA, GULP, GDIS
Engineering	Computational Fluid Dynamics, Modelling, Image Processing Tools: Matlab, OpenFoam

In a similar manner, a succinct overview of the research fields and their use of eScience techniques can be developed. This would include which techniques are used in which Research Phases, pros and cons of the techniques and the underlying algorithms used.From this a survey of the eScience practice can develop.

4.4. Characterising Research Projects

Another use for the BoK is to be able to efficiently characterise a research project. This may be for planning and resourcing, or for communicating the approach taken for support, reporting or publication purposes. At an individual project level, the information about techniques used in the Research Phases (3-7) would be positioned with information about the problem (1) and the overall research methodology (2). Compiling this information would provide the overall project workflow as follows: 1) Research Problem, 2) Methodology, 3) Data Collection, 4) Experimentation, 5) Analysis, 6) Communication, 7) Archive/Share

An example from Material Science:

1: An investigation of how different density functional theory (DFT) functionals perform when they are used to examine the interaction between two molecules.

2: To assess how each DFT functional performs, calculations for each functional would be performed and then the resulting atomistic and electronic properties would be compared to each other or literature theoretical or experimental results.

3: Data collection for this type of problem can be built from information published in similar problems, or with other computational software, and then this information is transformed into the required format for the particular software code we have chosen to use.

4: The experimentation or computation, in this case, may directly utilise code suites such as SIESTA[21], or GULP[22] on high-performance super-computing facilities, or utilise similar software via a Grid submission tool.

5: The analysis may be done using a tool such as GDIS[23], a Grid-enabled application for for the display and manipulation of isolated molecules and periodic systems. GDIS also allows the following functions to be performed through other packages: model rendering, energy minimisation, morphology calculation, space group processing or view the periodic table. Additional analysis may utilise visualisation software including GNU Plot[24], Povray[25], VMD[26] and Materials Studio[27].

6: Communication of results is primarily through peer-reviewed journal articles in the relevant area. Within the research group, plots and animations may be exchanged, with visualisation session using 3D stereo facilities to enhance understanding of the results.

7a: At this point, data is not typically shared with other research groups, although some electronic journals allow additional (supplemental) data, such as the optimised coordinates of structures, or enhanced visualisation images to be downloaded by researchers to improve understanding of the research.

7b: Input files for calculations are typically stored temporarily on the particular supercomputer that is used and/or on the researcher's workstation, with all resulting data backed up to the iVEC petascale datastore.

5. Conclusion and Future Work

In this paper we have described the need for an eScience BoK, particularly as a resource for educators. The value does extends beyond education, as a means for researchers to survey applicable techniques in their problem space. Support for eScience can also utilise the BoK for infrastructure planning, staff development and communication of tangible aspects of eScience.

In the future, the BoK will be made available for contribution, discussion and evolution. The mechanism for this sharing should eventually support mash-ups for reusing the knowledge base. For example, the BoK could provide the back-end to a research recommender system. As eScience grows in usage and techniques, tools to support researchers in understanding and selecting techniques will be of value. Support and education of the researchers will be assisted by the availability of a comprehensive guide to technology - the eScience BoK.

Acknowledgment

Thanks to my colleagues at iVEC for feedback and discussion of the eScience BoK idea, and to the researchers who provide information on their projects.

About iVEC: Our mission is to foster and promote scientific and technological innovation, through the provision of Supercomputing and eResearch services to the research community, commercial organisations and government agencies. iVEC is a joint venture among CSIRO, Curtin University of Technology, Edith Cowan University, Murdoch University and The University of Western Australia and is supported by the Western Australian Government.

Bibliography

[1] T. Clark and J. Kinoshita *Alzforum and SWAN: the present and future of scientific web communities.* Briefings in Bioinformatics, vol 8(3) p163-171, 2007.

[2] National Nuclear Security Administration, "NNSA Administrator Addresses Next Generation of Computational Scientists", Department of Energy: Computational Science Graduate Fellowship Conference, `http://www.nnsa.energy.gov/mediaroom/speeches/csgfremarks062210`, Jun 22, 2010

---

[21] `http://www.icmab.es/siesta/`

[22] `http://www.ivec.org/gulp`

[23] GTK Display Interface for Structures: `http://gdis.sourceforge.net/`

[24] `http://www.gnuplot.info/`

[25] Persistence of Vision Raytracer: `http://www.povray.org/`

[26] Visual Molecular Dynamics:`http://www.ks.uiuc.edu/Research/vmd/`

[27] `http://accelrys.com/products/materials-studio/`

[3] T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 2009.
[4] H. Foreman, “Editors’ Update: Your network for knowledge”, Elsevier Editors’ Update, Issue 30, June 2010.
[5] R. Sidi, “The Heart of a New Science Ecosystem”, Elsevier Editors’ Update, Issue 30, June 2010.
[6] B. Mann, M. Dove, M. Mineter, J. Oliver and R. Sinnott, “Education and Training in UK e-Science”, UK e-Science Technical Report Series UKeS-2005-01, URL: http://www.nesc.ac.uk/technical_papers/UKeS-2005-01.pdf
[7] ICEAGE, Report from the ICEAGE Curricula Development Workshop, Brussels, 14-15 February 2008
[8] G.F. Fox, “Parallel Computing and Education”, Daedalus, Vol 121(1), Winter 1992, p111-118.
[9] P.D. Lax, “Report of the Panel on Large Scale Computing in SCience and Engineering”, Sponsored by Department of Defence and the National Science FOundation, December 1982, Accessed March 2011, URL: http://www.pnl.gov/scales/docs/lax_report_1982.pdf
[10] M. Anderson and E. Vander Meer, “Policy for Supporting Grid and e-Science Education and Training”, GFD-I.153, Open Grid Forum, 2009.
[11] S.C. Glotzer (Chair) et.al., International Assessment of Research and Development in Simulation-Based Engineering and Science, World Technology Evaluation Center, Inc. 2009
[12] T. I. Oren, *Toward the Body of Knowledge of Modeling and Simulation*, Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2005.
[13] Institute of Electrical and Electronics Engineers, Inc., Guide to the Software Engineering Body of Knowledge, New York, NY, USA. 2004. Available online http://www.swebok.org.
[14] B. Pink and G. Bascand, “Australian and New Zealand Standard Research Classification (ANZSRC)”, ABS Catalogue No. 1297.0, 2008.
[15] V. Maxville, ”Preparing scientists for scalable software development,” SECSE, pp.80-85, 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering, 2009
[16] D.J. Meder, V. Pankratius and W.F. Tichy, “Parallelism in Curricula - An International Survey”, The International Working Group on Software Engineering for Parallel Systems (SEPARS), December, 2009.
[17] M.J. Eppler, “A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing”, Information Visualization, Vol 5, p.202 –210, 2006.
[18] S. Bresciani and M.J. Eppler, “The Benefits of Synchronous Collaborative Information Visualization: Evidence from an Experimental Evaluation”, IEEE Transactions on Visualization and Computer Graphics, Vol. 15 (6), November/December 2009.
[19] P. Navratil, “Overview and Introduction to Scientific Visualization”, Scaling to Petascale Summer School, July 9, 2010.
[20] J. Decrane and T. Hinze, “A Multidisciplinary Survey of Computational Techniques for the Modelling, Simulation and Analysis of Biochemical Networks”, Journal of Universal Computer Science, vol. 16, no. 9 (2010), 1152-1175.
[21] J.B. Wang, C. Hines and R. D. Muhandiramge, “Electronic structure of quantum dots and other nanosystems”, invited review chapter in the Handbook of Theoretical and Computational Nanotechnology, vol 10, chapter 10, p545-604, American Scientific Publishers, 2006.